

Comparative genomics and population resequencing elucidate streamlined immune adaptations to aquatic environments in cetaceans

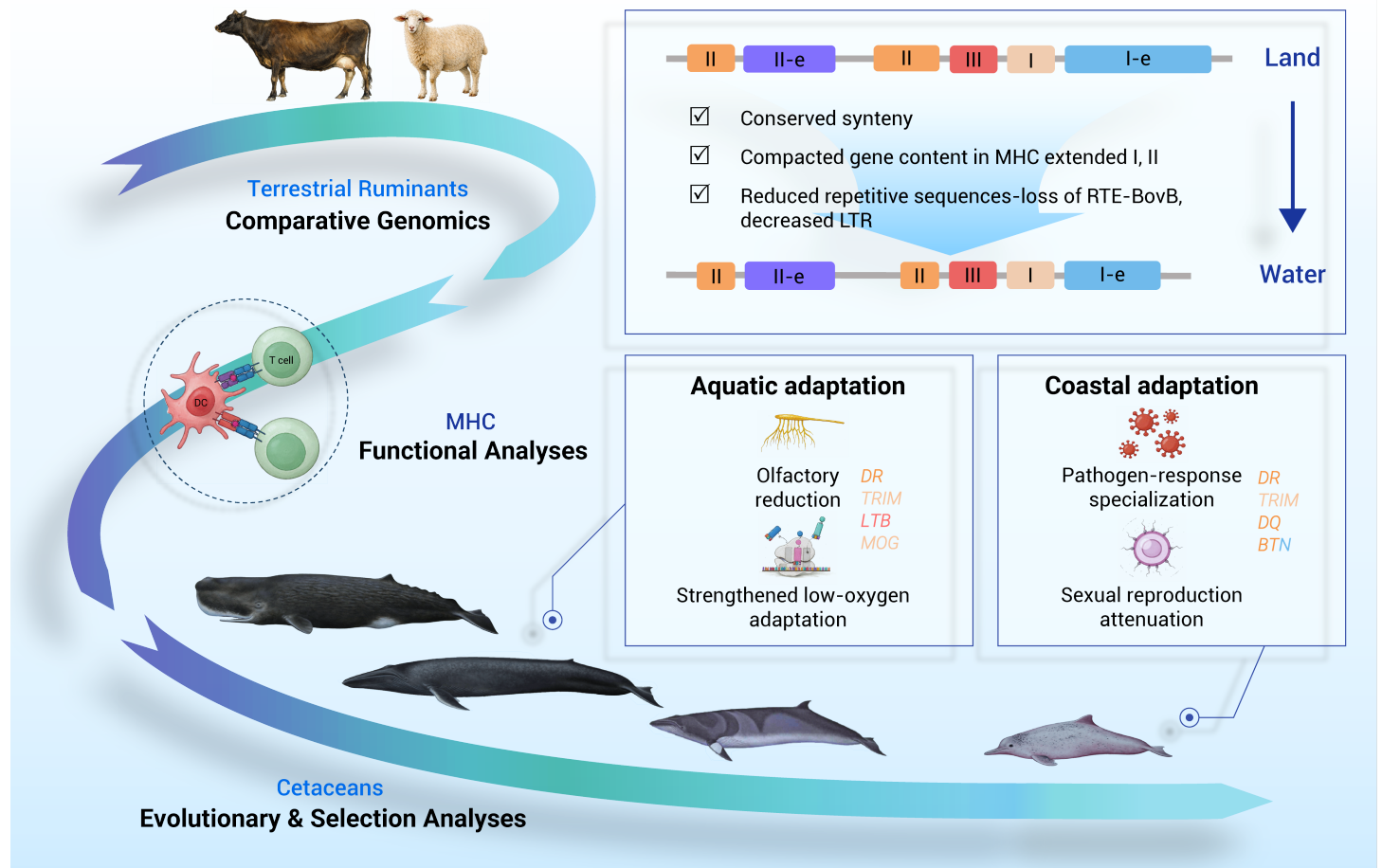
Dongjiao Liu,^{1,2,5} Jun Liu,^{1,2,5} Peiyu Xie,^{1,2} Liyuan Zhao,^{3,4} Hui Kang,¹ Qianhui Zeng,^{3,4} Songhai Li,¹ Xianyan Wang,^{3,*} and Peijun Zhang^{1,*}

*Correspondence: pjzhang@idsse.ac.cn (P.Z.); wangxianyan@tio.org.cn (X.W.)

Received: January 17, 2026; Accepted: March 11, 2026; Published Online: March 30, 2026; <https://doi.org/10.59717/j.xinn-life.2026.100245>

© 2026 The Author(s). This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- Cetaceans possess streamlined MHC regions with reduced gene content and repetitive elements.
- Cetaceans maintain conserved MHC structural frameworks.
- Class II genes mediated ancestral cetaceans' aquatic transition, Class I genes drove species diversity.
- Cetaceans exhibit aquatic adaptations, including olfactory reduction and metabolic downregulation.
- *Sousa chinensis* exhibit specialized pathogen defense mechanisms and attenuated reproductive gene sets.

Comparative genomics and population resequencing elucidate streamlined immune adaptations to aquatic environments in cetaceans

Dongjiao Liu,^{1,2,5} Jun Liu,^{1,2,5} Peiyu Xie,^{1,2} Liyuan Zhao,^{3,4} Hui Kang,¹ Qianhui Zeng,^{3,4} Songhai Li,¹ Xianyan Wang,^{3,*} and Peijun Zhang^{1,*}

¹Marine Mammal and Marine Bioacoustics Laboratory, Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya 572000, China

²University of Chinese Academy of Sciences, Beijing 101408, China

³Key Laboratory of Marine Ecological Conservation and Restoration, Ministry of Natural Resources, Xiamen 361005, China

⁴Fujian Provincial Key Laboratory of Marine Ecological Conservation and Restoration, Xiamen 361005, China

⁵These authors contributed equally to this work

*Correspondence: pjzhang@idsse.ac.cn (P.Z.); wangxianyan@tio.org.cn (X.W.)

Received: January 17, 2026; Accepted: March 11, 2026; Published Online: March 30, 2026; <https://doi.org/10.59717/j.xinn-life.2026.100245>

© 2026 The Author(s). This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

Citation: Liu D., Liu J., Xie P., et al. (2026). Comparative genomics and population resequencing elucidate streamlined immune adaptations to aquatic environments in cetaceans. *The Innovation Life* 4:100245.

Cetaceans have evolved numerous adaptations to thrive in aquatic environments. In this study, we identified and analyzed the entire MHC regions across cetacean species/populations and their terrestrial counterparts to elucidate immune adaptation strategies to aquatic life. Results reveal cetaceans have evolved a structurally streamlined MHC architecture, marked by reduced repetitive sequences (e.g., loss of BovB, decreased LTRs) and compacted gene content (class II and extended class I regions). Contracted gene sets were involved in olfactory transduction and reproduction. Class II genes were found critical for the initial aquatic transition in ancestral cetaceans, whereas class I genes facilitated subsequent habitat specialization and diversification. Strong signatures of positive selection were detected at both species level (e.g., in *DR*, *TRIM*, *LTB*, *MOG* genes) and population level (e.g., in *DR*, *TRIM*, *BTN*, *DQ* genes) within *S. chinensis*. These findings highlight the pivotal role of MHC-driven adaptations in enabling cetaceans to colonize and thrive in diverse aquatic niches. Collectively, our study provides key insights into the specialized immune evolution underpinning cetacean success in marine environments.

INTRODUCTION

Cetaceans are a group of marine mammals that evolved from terrestrial ancestors to become fully marine residents, a transition occurring approximately 55 million years ago.^{1,2} As apex predators, they play pivotal roles in marine ecosystems and serve as critical indicator species for ocean health.^{3,4} We know that the Paleocene-Eocene Thermal Maximum event drove cetacean ancestor back to the sea,² but the mechanisms by which their lineage flourished in the marine environment remain poorly investigated to date. A previous study demonstrated that high-performance metabolic and ecological innovations predominantly originated on land after the Ordovician, once organisms had overcome the challenges of life away from water, indicating that marine ecosystems may be more stable and hospitable than terrestrial ones, and therefore maintained reduced selective pressures.⁵ Huelsmann et al. also demonstrated that the loss of some genes during the land-to-water transition in cetaceans contributed to genomic changes associated with aquatic adaptations.⁶ Other adaptations, including specialized reproductive strategies, thermoregulatory mechanisms, cancer resistance, and advanced biosonar systems, have also been reported.^{1,7-9} Notably, their immune system is believed to have undergone significant evolutionary modifications during the shift from land to sea, enabling survival and diversification across diverse marine habitats.¹⁰

The major histocompatibility complex (MHC) is a core component of the immune system, found exclusively in jawed vertebrates.¹¹ MHC genes encode glycoproteins that are essential for the self/non-self discrimination critical to immune responses.¹²⁻¹⁴ They are classified into three main categories: class I, class II, and class III. Class I molecules are ubiquitously expressed on nucleated cells, presenting endogenous antigens (e.g., from virus-infected or cancerous cells) to CD8⁺ T cells. Class II molecules, restricted to professional antigen-presenting cells (APCs) such as dendritic cells, macrophages, and B cells, present exogenous peptides to CD4⁺ T cells.¹⁵ Class III genes encode secreted complement factors (e.g., C2, C4, and B) involved in immune responses.^{14,16}

In the oldest jawed vertebrates, MHC classes I, II, and III are typically linked.¹⁷⁻²¹ Subsequent gnathostome ancestors exhibited a rearranged sequence of class II, I, and III genes, with class Ib genes remaining on the same chromosome but outside the core MHC region.¹¹ Marsupials display an interspersed organization of class I and II genes, whereas placental mammals retain a relatively conserved MHC structure, with class I, class III, and class II regions arranged in a defined order.²² Strikingly, cetaceans possess a distinct MHC class II region characterized by two discontinuous subregions (class IIa and IIb), a result of an autosomal inversion.^{23,24} However, the evolutionary significance of this inversion remains unresolved.

The MHC is among the most polymorphic regions in vertebrate genomes, primarily driven by pathogen-mediated selection that maintains genetic diversity.^{25,26} Its evolution follows a complex “birth-death” model, involving gene duplication, diversification through ectopic recombination, and pseudogenization.²⁷ This dynamic process generates high copy number variation, further complicating analyses of genetic diversity and phylogenetic relationships within the MHC.

As a clade of “secondarily aquatic” mammals, cetaceans represent a unique evolutionary lineage and serve as a model for studying molecular mechanisms of environmental adaptation.² The Indo-Pacific humpback dolphin (*Sousa chinensis*), a subtropical-to-tropical inshore species with high habitat fidelity, inhabits ecologically distinct coastal shallow waters.²⁸ Given escalating threats from global environmental changes and anthropogenic activities, *S. chinensis* emerges as an ideal species for investigating cetacean immune system evolution and adaptation.²⁹

In this study, we construct high-resolution maps of the MHC structures across cetacean species and their terrestrial relatives, using human MHC as an outgroup. Through comparative genomics, we explore immune system evolution at genetic and genomic levels, focusing on the cetartiodactyl transition from land to sea. By analyzing MHC gene divergence, selective pressures, and evolutionary dynamics, we aim to elucidate their role in environmental adaptation. Furthermore, we examine population-level genetic diversity and evolutionary patterns in several *S. chinensis* populations from the South China Sea, providing insights into their immune system adaptation to aquatic environments. Our findings will establish a foundational understanding of cetacean immune mechanisms and inform conservation strategies for this ecologically and evolutionarily significant group.

MATERIALS AND METHODS

MHC annotation

The raw genome sequencing data of the *S. chinensis* were obtained from our previous study.³⁰ Tandem repeats were predicted using Tandem Repeat Finder v4.09. An ab initio repeat library was constructed with RepeatModeler v2.0.4.³¹ Long terminal repeats (LTRs) were annotated using LTR_retriever v2.9.9,³² and the annotated LTR sequences were merged with de novo predicted results, excluding the de novo annotated LTR library. RepeatMasker v4.1.2,³³ combined with Dfam3.7 and RepBase v18.10, as well as the ab initio repeat library, was used to identify homologous and novel repeats.

Gene annotation involved two approaches. Ab initio annotation was conducted using AUGUSTUS v3.4.0,³⁴ while homology-based annotation utilized well-annotated protein sequences downloaded from NCBI for species

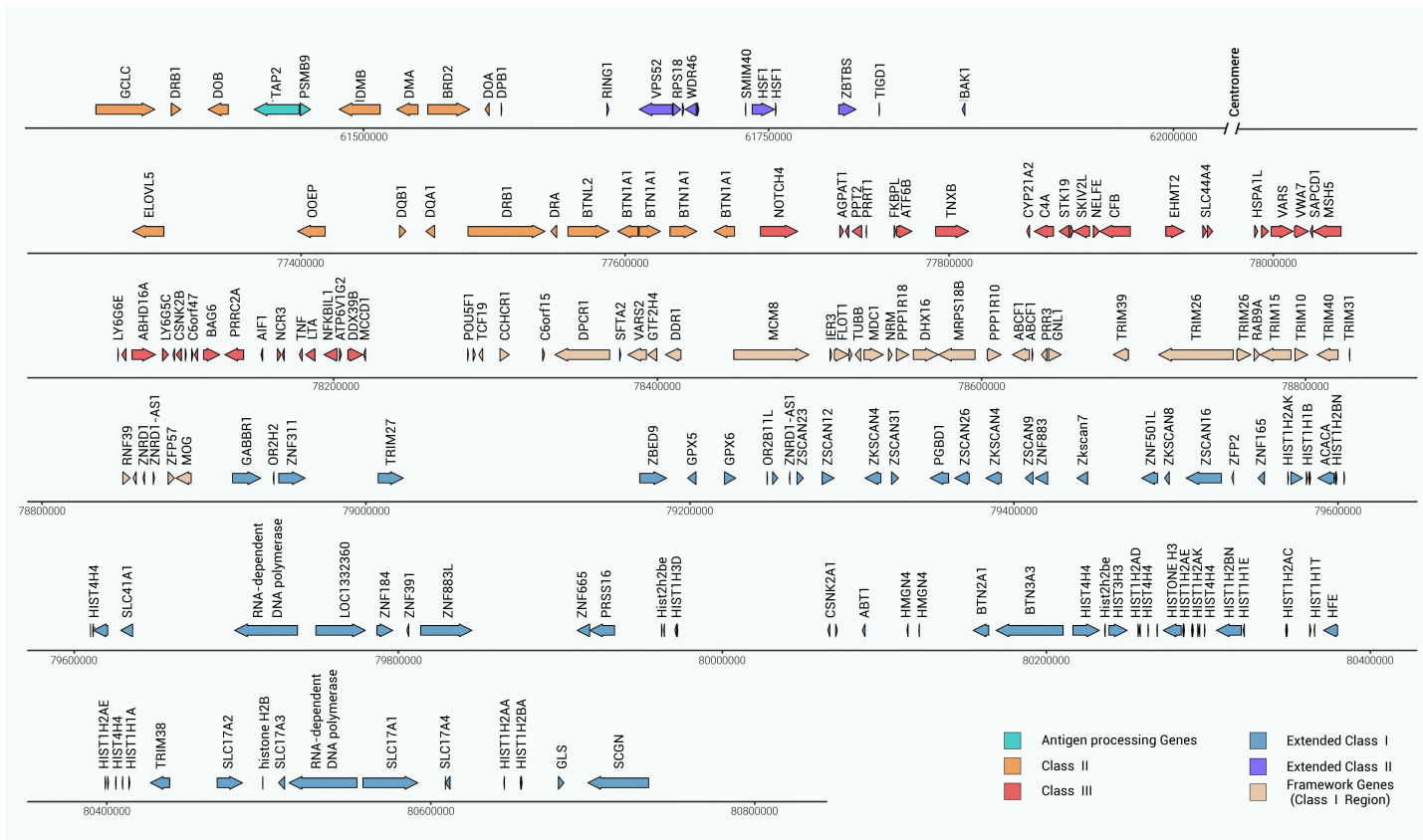


Figure 1. MHC structure of the *S. chinensis*.

such as human (*Homo sapiens*), cattle (*Bos taurus*), sheep (*Ovis aries*), bottlenose dolphin (*Tursiops truncatus*), blue whale (*Balaenoptera musculus*), sperm whale (*Physeter macrocephalus*), Yangtze finless porpoise, minke whale (*Balaenoptera acutorostrata*), and killer whale (*Orcinus orca*). Additionally, HLA gene protein sequences were retrieved from the HLA Nomenclature website (<https://hla.alleles.org/alleles/index.html>). Protein sequences with high-quality annotations were also obtained from the UniProt database using gene names as search criteria, and these sequences were incorporated into the GeneWise v2.4.1 database.³⁵ The results of gene prediction were integrated using EvidenceModeler v1.1.1³⁶ and subsequently validated through comparison with a more comprehensive protein database using the online NCBI BLAST system, yielding the final annotation. Using the same methodology, we identified the complete MHC sequences of cattle, sheep, blue whale, sperm whale, and minke whale as well.

Multi-species gene collinearity analysis and confirmation of MHC region

We selected the following species for multi-species gene collinearity analysis: *S. chinensis*, *B. musculus*, *P. macrocephalus*, *B. acutorostrata*, and their terrestrial relatives *B. taurus* and *O. aries*, as well as *H. sapiens*, all of which possess high-quality chromosome-level assembled genomes. Genome data for these species were obtained from NCBI (Table S1).

Pairwise comparisons between species were conducted using DIAMOND (v2.0.11),³⁷ and the results were analyzed with MCScanX to identify homologous segments across species.³⁸ Gene clustering analysis and the identification of single-copy orthologous genes were performed using OrthoFinder (v2.5.5) with a default E-value threshold of 0.001.³⁹ All identified gene sequences were named according to the vertebrate MHC nomenclature rules,⁴⁰ such as the *S. chinensis* MHC gene named Soch_MHC.

Based on the results of gene clustering and functional annotation, custom R (v4.3.1) scripts were used to generate the MHC gene structure map for the *S. chinensis*. Collinearity maps between species were created using NGenomeSyn (v1.41) scripts, with orthologous genes explicitly annotated.

Divergence time of MHC

Based on the MHC genomic sequences of *H. sapiens* (as the outgroup), *S.*

chinensis, *B. musculus*, *P. macrocephalus*, *B. acutorostrata*, *B. taurus* and *O. aries*, phylogenetic analysis was conducted using the identified single-copy orthologous genes. The coding sequences (CDS) were concatenated to form inferred sequences for phylogenetic reconstruction. Phylogenetic trees for MHC class I, II, and III genes, the entire MHC region, as well as for whole genome sequences were constructed using RAXML (v8.0.26),⁴¹ and the results were visualized for presentation.

The molecular clock theory posits that the rate of substitution of codons in a species' gene sequences is largely unaffected by time, suggesting a relatively constant evolutionary rate. Therefore, the genetic distance between two species is proportional to their divergence time.⁴² In this study, we referenced the divergence times of species from the TimeTree database (<https://time-tree.org>), alongside fossil record data, as calibration points. The basic information of divergence time correction point is shown in Table S2. Finally, using the mcmctree module in PAML v4.10.7,⁴³ we estimated the divergence times of different MHC genes by analyzing the whole MHC region, as well as the MHC class I, II, and III gene regions for seven species.

Other methods

Methods relating to statistics, MHC expansion and contraction, positive selection, resequencing data processing, population polymorphism analysis, and population environmental adaptation analysis are provided in the supplementary file.

RESULTS AND DISCUSSION

Structure of the MHC region in the *S. chinensis*

Through genome annotation and multi-species gene collinearity analysis, we successfully isolated and analyzed the intact MHC regions of six species, *S. chinensis*, *B. musculus*, *P. macrocephalus*, *B. acutorostrata*, *B. taurus* and *O. aries*, while using *H. sapiens* MHC as the outgroup. Here we further showcase the MHC details of *S. chinensis* (Figure 1, please see MHC structures of the other 3 cetacean species in Figures S1–S3). The MHC region of the *S. chinensis* is bounded by one of the flanking markers SCGN and GCLC on the other end, spans 3.97 Mb, and contains 219 genes. Please see data details in Table S3. As shown in Figure 1, the MHC region of the *S. chinensis* consists

Table 1. Distribution of MHC genes in the *S. chinensis*.

Subregion	Length (Mb)	Percentage (%)	Gene Count	Gene density (Gene count/Length)
I	0.62	16.00	39	62.90
II	0.68	17.00	24	35.29
III	0.63	16.00	46	73.02
I -extended	1.82	46.00	98	53.85
II -extended	0.22	5.00	12	54.55
Total	3.97	100.00	219	55.16

of five subregions: II, II-extended, III, I, and I-extended. The II-extended region is linked to the class II region, and the class I region is connected to the I-extended region, with the class III genes in between. The gene categories and their arrangement are highly similar to the published human HLA region genes.⁴⁴ Notably, although we did not annotate any classic class I genes (i.e., *A*, *B*, *C*) in the *S. chinensis* MHC region, we still categorized the framework genes into the class I subregion, following the convention established in the HLA region,⁴⁴ where classic class I genes are interspersed with framework genes.

MHC class II gene region is interrupted by the centromere on chromosome 10. The intervening sequence is about 15.40 Mb long, dividing the MHC class II gene region into two subregions, IIa and IIb. The IIa segment includes classic and non-classic gene loci such as *DOA*, *DOB*, *DMA*, *DMB*, and *DPB*, while the IIb segment contains classic gene loci such as *DRA*, *DRB*, *DQA*, and *DQB*. The II-extended region genes are inserted between the IIa and IIb gene regions, differing from the relationship between HLA class II and II-extended genes in humans.⁴⁴

The distribution of class I, II, and III gene regions is shown in Table 1. The MHC class I and related I-extended gene region are the longest, with a total length of approximately 2.44 Mb and the highest number of genes, 137. The class II and related genes follow, with a length of 0.90 Mb and 36 genes. The class III gene region is the shortest, it has 46 genes, indicating a higher gene distribution density compared to other subregions.

In ancestral mammals, class I and class II genes are interspersed, and the distribution sequence of MHC gene subregions is extended class II, class I/II, class III, framework, extended class I.²² With the evolution of mammals, such as in mice, class I genes have interspersed across class III genes and framework genes, a pattern also observed in marine mammals. Before the speciation between Artiodactyla and Cetacea, class I genes underwent reorganization across class III genes, and subsequently, frequent contractions and expansions occurred in different species.²² Specifically, Fukami-Kobayashi et al. (2005) suggest that long interspersed nuclear element (LINE) sequences can trigger the duplication of genomic fragments, generating paired duplicated genomic segments, which may lead to the duplication of class I genes within framework genes.⁴⁵ The linkage between class I genes and framework genes may impact the polymorphism of class I genes and their co-evolution, closely related to the expansion and contraction of the class I gene family.²²

Comparative analysis of repetitive sequences in representative cetaceans and their terrestrial relatives

To examine the differences of MHC genes between cetaceans and their terrestrial relatives at genomic level, we conducted a repetitive sequence (i.e., TE elements, Tandem Repeat) analysis. Compared to the genome-wide content, the proportion of TE elements in MHC region in *B. taurus* and *O. aries* is not significant ($P > 0.05$). In *B. taurus*, the repetitive sequence content in the MHC region is 46.27%, while in the entire genome, it is 46.23%. For *O. aries*, the repetitive sequence content in the MHC region slightly decreases from genome-wide 43.66% to 41.72%. However, in cetaceans, the proportion of repetitive sequences in the MHC region slightly decreases: *S. chinensis* from 35.42% to 28.68%, *B. acutorostrata* from 44.28% to 33.72%, *B. musculus* from 36.10% to 33.46%, and *P. macrocephalus* from 34.93% to 30.69%, indicating that MHC functions in cetaceans were probably streamlined or specialized compared to their terrestrial relatives ($P > 0.05$, Table S4).

Specifically, the content and composition of Retroelements are similar to

each other both within cetaceans and within their terrestrial relatives, but differs largely across groups (Table S4). Retroelements reduce significantly in cetaceans ($P < 0.05$, Table S5), which may primarily due to the loss of RTE elements and LTR elements (Figure 2 & Table S4). Ruminants such as cattle and sheep have a highly expanded RTE-BovB type element, accounting for about 17-19% of their genomes and approximately 16% in the MHC region. In cetaceans, however, the RTE-BovB type element is almost all lost, both in MHC and whole genome (Figure 2 & Table S4). Besides, we also find that LTR elements in cetaceans are significantly reduced compared to their terrestrial relatives ($P < 0.05$), their proportion in whole genome and in MHC regions is relatively the same (Figure 2 & Table S5). Our results indicate that the loss of RTE-BovB and LTR in cetaceans is likely an adaptation strategy to their aquatic environment.

Though the LINEs in cetaceans are significantly reduced, it is because they nearly lost all the RTE-BovB ($P < 0.05$, Table S5). The L1/CIN4 in cetaceans are obviously expanded, and the proportions of L1/CIN4 in cetacean genomes are higher than in their MHC genes (Figure 2 & Table S4), indicating a positive adaptation to the aquatic environment, but with limited relevance to immune adaptation.

Also, cetaceans have significantly higher proportions of SINE elements ($P < 0.05$, Table S5 & Figure 2), with a higher distribution in MHC genes than in genomes (Table S4 & Figure 2), indicating that these elements are important for their immunity in aquatic environment, as well as contribute to adaptation to the aquatic environment.

Compared to their terrestrial relatives, cetaceans have a lower proportion of repetitive sequences in the MHC region, which may be a specific result of gene contraction in this region. Our study found that cetaceans have at least 8% fewer repetitive sequences in their genomes compared to their terrestrial relatives. A certain study examined 11 cetacean species, 6 pinniped species, and their terrestrial relatives, finding that most marine mammals had about 10% fewer repetitive sequences than their terrestrial relatives, with similar compositions of repetitive elements at the DNA transposons, LINEs, LTR elements, and SINEs levels. In contrast, ruminants have a richer repertoire of five types of repetitive sequences, including LINE/RTE-BovB, LTR/ERV1, LTR/ERVK, SINE/CORE-RTE, and SINE/tRNA-CORE-RTE.⁸ L1 and L2 both belong to LINEs. Our study found that the content of L1 and L2 in cetaceans is higher than in their terrestrial relatives, both at the MHC and genomic levels (Figure 2). L1 has autonomous transposition ability, can self-replicate, and inserts into the genome through a reverse transcription mechanism. L1 elements encode two proteins, ORF1p and ORF2p, which have RNA-binding and reverse transcriptase activities, respectively. These LINEs can influence genomic stability and gene expression regulation in cetaceans, playing a crucial role in their genomic evolution and adaptation.⁴⁶

Conservation and contraction of MHC regions in cetacean

In addition to repetitive sequences, we further analyzed the genes of MHC regions. The MHC loci in these seven mammals are tightly linked within specific chromosomal regions. The arrangement of MHC genes across the seven species shows a high degree of linear relationship. In comparison to *H. sapiens*, where class II genes form a continuous region, the class II region in *B. taurus* is divided into two subregions, IIa and IIb, separated by the centromere. The gene arrangement in the II-extended region and the IIa region is reversed compared to that in humans. While minor segmental inversions of class II and II-extended genes are observed among *H. sapiens*, *B.*

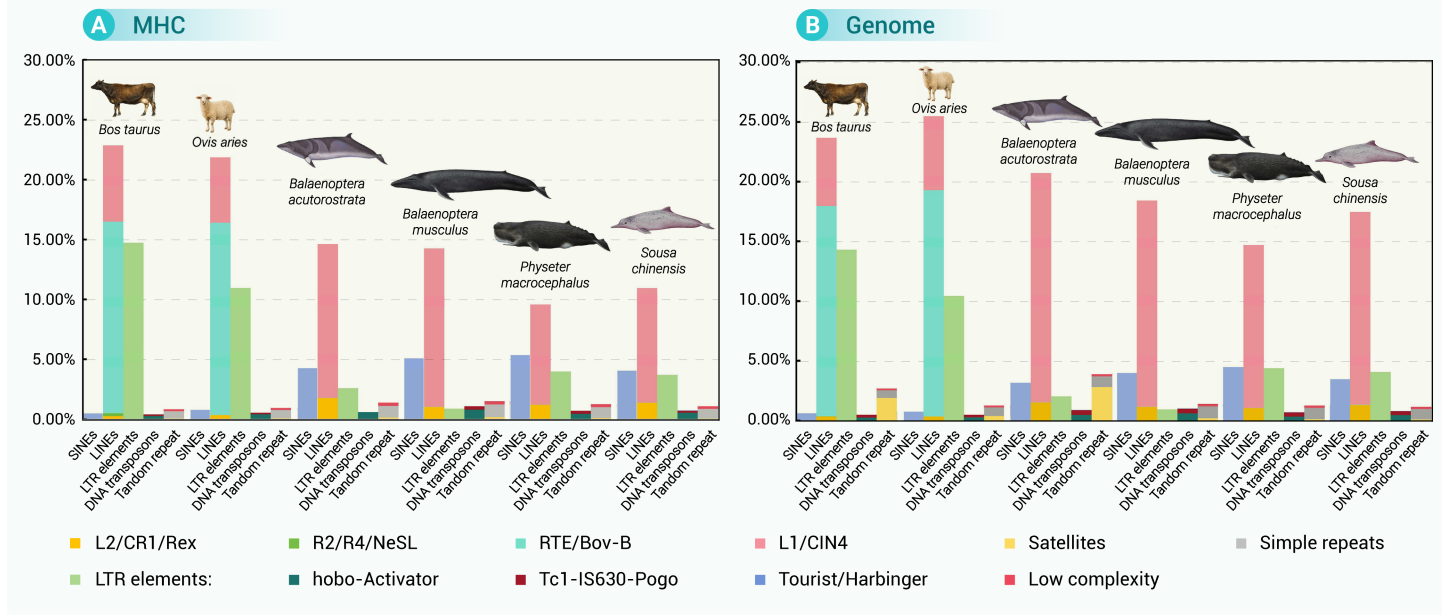


Figure 2. Ratio of repetitive elements in representative mammalian species (A) depicts the content and composition of repetitive elements within the MHC region. (B) illustrates the content and composition of repetitive sequences across the entire genome. Five major repetitive element categories are represented: SINEs, LINEs, LTR elements, DNA transposons, and Tandem repeats. A detailed breakdown of subcategories is provided in the key at the bottom of the figure.

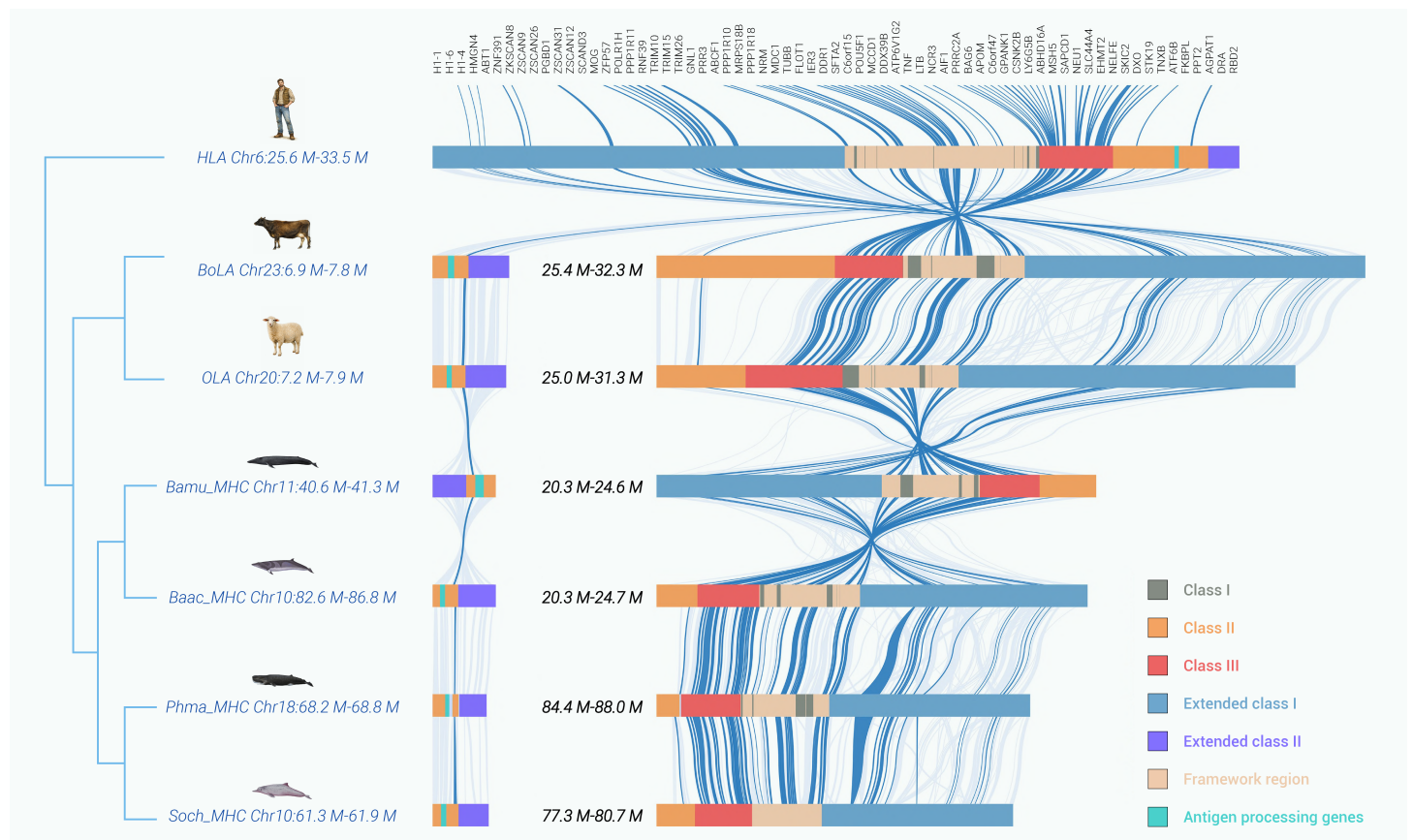


Figure 3. Synteny of MHC gene structure cross multiple species The connecting lines between chromosomes of different species illustrate homologous genes across species. Dark blue lines represent single-copy orthologous genes. Gene names are listed at the top of the figure. Species are arranged from top to bottom as follows: *H. sapiens*, *B. taurus*, *O. aries*, *B. musculus*, *B. acutorostrata*, *P. macrocephalus*, *S. chinensis*.

taurus, and *O. aries*, the overall linear arrangement of the remaining genes is well conserved. Notably, compared to human, cetaceans exhibit high conservation and significant linearity in the class II and II-extended regions ($P < 0.05$), with no inversions detected across *B. taurus* and *O. aries* (Figure 3).

The comparative analysis of the MHC regions between cetaceans and their

terrestrial relatives reveals distinct evolutionary modifications in cetaceans. Compared to terrestrial mammals, the MHC regions in cetaceans appear more compact, with reduced gene content, especially in the extended class I and class II regions, potentially reflecting a streamlined immune system adapted to aquatic environments. Despite the reduction in gene content,

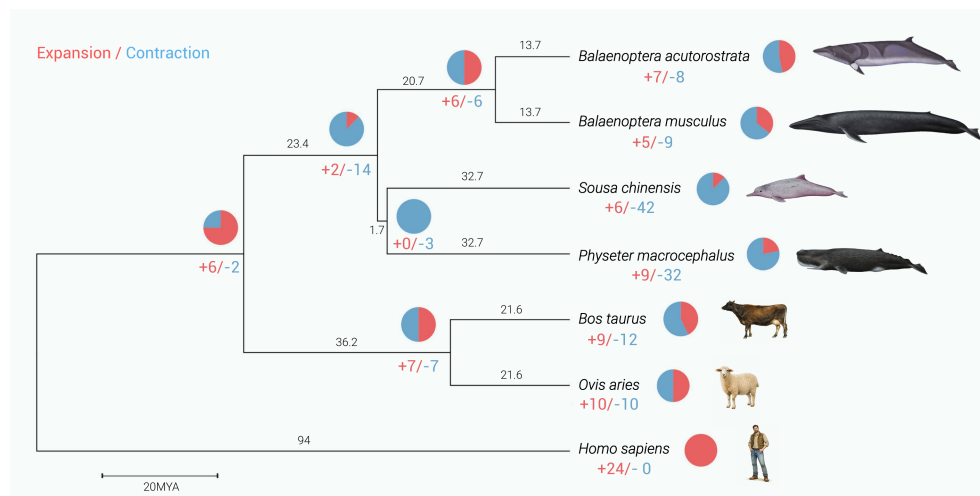


Figure 4. The expansion and contraction of MHC orthogroups in 7 species.

show contraction, whereas 14 orthogroups show expansion, specifically, seven orthogroups exhibit significant contraction ($P < 0.05$, Table S7 & Figure 4). Compared to the other three cetacean species, the *S. chinensis* does not display any significant orthogroups undergoing expansion or contraction ($P > 0.05$). However, it does exhibit contraction in 38 orthogroups and expansion in 17 orthogroups, with two of these orthogroups showing more than three instances of expansion (Table S7). These two orthogroups primarily consist of zinc finger proteins and histone genes, which may play crucial roles in neural development,

as well as in gene expression regulation and DNA repair.^{47,48}

cetaceans maintain the basic synteny and framework regions of the MHC. This suggests that while gene loss or pseudogenization occurred, the overall structural organization remains conserved. The antigen-processing genes remain conserved across species, indicating their critical role in immune function has been retained throughout cetacean evolution.

Gene family clustering identified 67 single-copy orthologous genes among the seven species (as listed in the upper side of Figure 3). They are 2 class II genes (*DRA* and *RBD2*), a large number of class III genes (29/74), framework genes (30/79), and class I extended region genes (13/79). These conserved regions can provide a reference for future identification of MHC regions in more mammalian species. Framework genes, including loci such as *MOG*, *PPP1R11*, *TRIM26*, *TRIM39*, *GNL1*, *POU5F1*, and *BAT1*, generally exhibit conservation within the MHC region across animals. Among all the genes, class III genes exhibit the highest level of conservation, accounting for 43.3% of all conserved orthologous genes, while class II genes show the lowest conservation. This suggests that class II and I genes, compared to III genes, have greater variability.

Statistics of the length and number of genes in each MHC subregion across the seven species are summarized in Table S6. The overall length of the MHC region varies among species, with terrestrial animals (6–8 Mb) exhibiting significantly longer lengths than cetaceans (~4–5 Mb) ($P < 0.05$). Calculations also reveal significant differences in gene density within the MHC regions among species ($P < 0.05$). The gene density in human HLA region is the lowest, approximately 33 genes/Mb, while *B. taurus* and *O. aries* MHC regions show similar densities, about 45–50 genes/Mb. In contrast, the density in cetaceans is significantly higher, about 50–60 genes/Mb. However, this phenomenon is due to the significantly higher gene density in class II region (30–42 genes/Mb) compared to terrestrial mammals (20–25 genes/Mb); for the other regions, gene density is comparable (Table S6). Class III gene density is the highest in all the investigated seven species and does not show clear aquatic adaptation traits. This phenomenon of reduced MHC region length and increased gene density in cetaceans likely reflect adaptations to a marine lifestyle, including different pathogen exposures and immune requirements, and they highlight the balance between maintaining critical immune functions and streamlining genetic regions.

MHC gene family expansion, contraction, and functional enrichment analysis

As we know that MHC regions in cetaceans are largely contracted for the adaptation of aquatic environment, except for the loss of RTE-BovB elements, we still do not know if there is contraction/expansion of important genes. MHC gene family expansion and contraction analysis shows that compared to terrestrial species, cetaceans exhibit a clear contraction of their MHC genes, with 14 orthogroups undergoing contraction and only 2 orthogroups undergoing expansion (Figure 4). In comparison to their common ancestor, the *S. chinensis* shows a more pronounced pattern of expansion and contraction, with 42 orthogroups undergoing contraction and only 6 orthogroups undergoing expansion.

Compared to terrestrial mammals, 87 MHC orthogroups in cetaceans

as well as in gene expression regulation and DNA repair.^{47,48}

We further conducted enrichment analysis for these orthogroups. To adapt to a specific environment, species are likely to undergo gene duplications and losses, leading to the expansion and contraction of genes.^{49,50} Previous studies on convergent evolution across species suggest that when species face selective pressures such as environmental changes or feeding competition, genes undergo directional expansions and contractions related to these pressures.⁵¹ In our present study, cetaceans show gene expansions and contractions associated with aquatic adaptations compared to terrestrial species. GO enrichment analysis of the expanded gene set in the cetaceans reveals enrichment in terms related to the negative regulation of nitrogen compound metabolism (GO:0051172), negative regulation of macromolecule biosynthetic processes (GO:0010558), and negative regulation of cellular metabolic processes (GO:0031323) (Table S8 & Figure S4). These genes may be associated with adaptation to low-oxygen conditions in the aquatic environment. A study on the brain transcriptome of the harbor seal showed that compared to its terrestrial relative, the ferret, genes related to oxidative phosphorylation and other high-energy-consuming processes were significantly downregulated. This metabolic reduction helps the seal survive in hypoxic conditions by conserving oxygen and reducing energy needs, thus extending dive times.⁵²

KEGG enrichment analysis of the contracted gene set in the cetaceans highlights pathways related to natural killer cell-mediated cytotoxicity (ko04650) and antigen processing and presentation (ko04612), which may be linked to adaptation to varying pathogen pressures in the aquatic environment (Table S8 & Figure S4). Pathogen types in aquatic environments differ from those on land in factors such as temperature, salinity, and oxygen levels. Some pathogens may not exist in aquatic environments or pose less of a threat to aquatic animals, which could reduce the selective pressure on certain immune pathways, leading to gene contraction. The contracted gene set also includes functions related to olfactory transduction (ko04740), which may be linked to adaptations in response to aquatic environments. A study on populations at different altitudes found that the most significant adaptive difference was a reduction in olfactory response in high-altitude regions.⁵³ Similarly, gene family contractions related to olfaction were found in foxes living at high altitudes, suggesting that extreme environment conditions are a key factor in the contraction of olfactory genes.⁵⁴ modern cetaceans have lost a substantial number of olfactory receptor genes.⁵⁵

To specifically investigate how defined cetacean species adapt to their unique habitat, we use *S. chinensis* as a representative, which is characterized as tropical/subtropical coastal species with high habitat fidelity.²⁸ The results reveal several adaptive traits in tropical/subtropical coastal waters (Figure S5 & Table S9). KEGG enrichment analysis of the expanded gene set suggests pathways related to necroptosis (ko04217) and neutrophil extracellular trap formation (ko04613), which may be linked to adaptation to greater pathogen pressures in coastal environments (Figure S5B). Neutrophils, the most abundant white blood cells in the blood, play a crucial role in host defense and chronic inflammatory diseases. They can release their DNA in the form of NETs, which trap bacteria. Coastal waters often face higher pathogen pressures than open-ocean regions due to the complexity of the

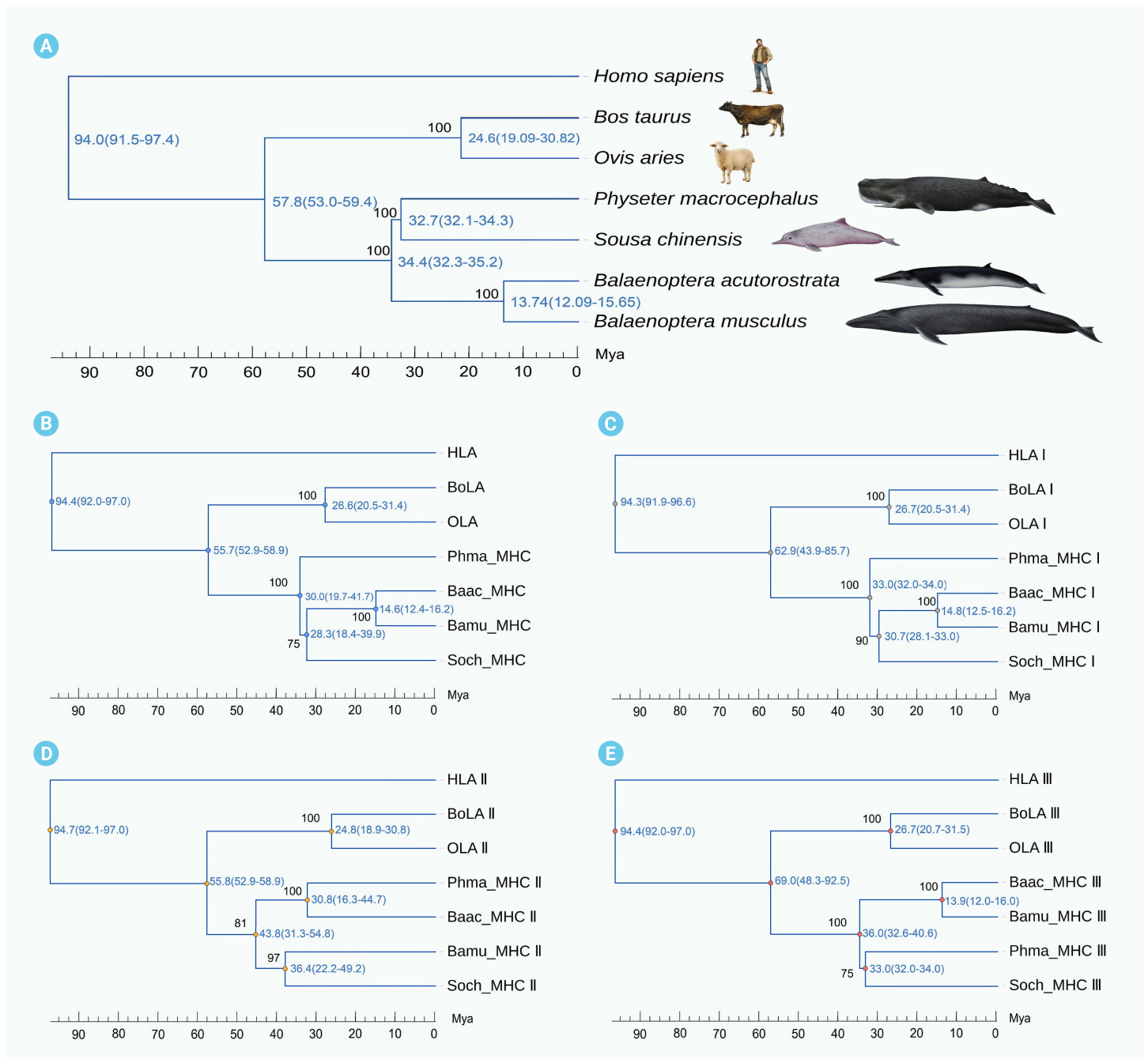


Figure 5. Phylogenetic tree and divergence times of the investigated species based on different sequences (A) to (E) represent the analysis results based on the whole genome, MHC regions, MHC class I region, class II region, and class III region, respectively. Blue data points indicate the divergence times at each node.

ecosystem and human activities such as agricultural runoff and pollution. NET formation may thus be a key mechanism for the *S. chinensis* to cope with coastal pathogen pressures. GO enrichment analysis of the contracted gene set reveals enrichment in terms related to reproduction, indicating that breeding at coastal waters may face less immune pressure comparing to oceanic or deep-sea waters (Figure S5C).

The divergence time of the MHC in cetaceans

To further investigate when different classes of MHC genes evolved during transition of ancestral cetaceans from land to sea, we reconstructed the phylogenetic tree of them. According to the data of our selected seven species, we find that phylogenetic tree based on whole genome data (Figure 5A) and MHC regions (Figure 5B) are different, indicating a different evolution pattern between them. To identify how different genes specifically evolved, we further constructed their phylogenetic tree based on different MHC subregions (Figures 5C-E). Results show that the topology change of the MHC trees are

due to changes in class I and II genes indicating a different evolution pattern with the whole genome, whereas the topology of class III tree are highly the same with that of whole genome data (Figure 5E). These results indicate that MHC class I and II genes play a vital role for the adaptation of water environment in cetaceans while the highly conserved class III genes show little correlation with adaptation to water environment. Besides, the divergence time of class I and III genes between ruminants and cetaceans was estimated to be 62.9 (95% confidence interval, 43.9-85.7) Mya, and 69.0 (48.3-92.5) Mya, respectively; while the divergence time for MHC class II genes was 55.8 (52.9-58.9) Mya. The divergence times of class I and III genes were earlier. This further supports the phenomenon of MHC class I gene migration, where class I genes crossed over and rearranged with class III genes, subsequently undergoing frequent expansions and contractions in different mammalian species.²²

When we look into the MHC gene tree regarding all the cetacean species, we find that for MHC class II genes (55.8 Mya), they split after the divergence

of ruminants and cetaceans 57.8 (53.0–59.4) Mya, but earlier than the following cetacean species diverse 34.4 (32.3–35.2) Mya (Figure 5D), suggesting that evolution of class II genes in cetaceans are critical for the fast adaptation of their ancestor to aquatic environment, but for the subsequent cetacean species diverse for different habitats, class II genes worked less. For class I genes, situation is different. These genes split before the cetacean ancestors entered the sea, 62.9 (43.9–85.7) Mya vs 57.8 (53.0–59.4) Mya, but evolved again during the subsequent split of cetacean species and caused a different phylogenetic tree topology compared to the whole genomic data. The evolution pattern of class I genes suggests they may be critical for the adaptation of different aquatic habitats, and thus is important for the diversity of cetaceans.

Positively selected MHC genes for aquatic adaptation

To further investigate the important MHC genes for aquatic environmental adaptation, we analyze which genes are positively selected. Compared to *B. taurus* and *O. aries*, cetaceans exhibit significant positive selection in 17 genes: *BRD2*, *DRA* from class II region, *ApoM*, *TNF*, *CSNK2B*, *ABHD16A*, *LTB*, *MCCD1* from class III region, *ZBED9*, *ZSCAN31*, *C6orf15*, *MRPS18B*, *MOG*, *TRIM15*, *PPP1R11*, *ZNRD1*, *PPP1R10* from class I region. Compared to three other cetacean species, *S. chinensis* shows significant positive selection in nine genes: *DRA* from class II, *LTB* from class III, *MOG*, *MRPS18B*, *PPP1R11*, *TRIM15*, *C6orf15*, *ABCF1*, *ZBED9* from class I. The specific positively selected amino acid sites are detailed in Table S10. Among them, eight repetitively detected genes, i.e., *DRA* from class II, *LTB* from class III, *ZBED9*, *C6orf15*, *MRPS18B*, *MOG*, *TRIM15*, and *PPP1R11* from class I, were found to be under positive selection throughout the entire cetacean evolution trajectory. The presence of positive selection in all these genes suggests that they have played critical roles in the aquatic adaptation of cetaceans and may have been especially important in enabling *S. chinensis* to thrive in its specialized nearshore and shallow-water environment.

Based on the results of gene family expansion and contraction, we find that during the aquatic transition of cetaceans, the gene family associated with *LTB* (Lymphotoxin Beta) underwent contraction, meanwhile the *LTB* gene itself was under positive selection. *LTB* encodes a type II membrane protein that functions through complex formation with lymphotoxin α , playing a key role in inflammation and immune regulation.^{56,57} The concurrent contraction of the gene family and positive selection on the retained copy suggests functional optimization, where redundant paralogs may have been lost and the remaining gene refined to carry out essential immune functions. In terrestrial environments, multiple *LTB*-related genes might be required to maintain a broader or more complex immune response, whereas in the aquatic environment, such complexity may no longer be necessary, and a single optimized copy may suffice.

In contrast, the *DRA*-related gene family showed expansion when modern cetaceans evolved to inhabit shallow coastal waters, and the *DRA* gene was also under positive selection. *DRA* encodes the α -chain of MHC class II molecules, which are crucial for presenting extracellular pathogen-derived peptide fragments to T cells, thus mediating adaptive immune responses (provided by RefSeq, 2020). The observed expansion of the *DRA* gene family may indicate an increased importance of this function in the marine environment, potentially requiring higher gene dosage or functional diversification to cope with pathogen exposure in coastal waters. Some of the duplicated copies may have undergone neofunctionalization or subfunctionalization, and positive selection could have driven the fine-tuning of these functions. This may reflect an enhanced immune defense mechanism in species such as the *S. chinensis*, which inhabits pathogen-rich nearshore environments.

Population polymorphism analysis

In the aforementioned sections, we discuss the evolution of MHC regions of mammals during the transition from land to the sea at species level. We also want to know how their MHC genes evolve at population level. Considering this, we take *S. chinensis*, a highly habitat faithful species as an example. We used five geographical representative *S. chinensis* populations data from the South China Sea: Leizhou Bay (LZB), Sanniang Bay (SNB), Jiangmen (JM), Quanzhou-Xiamen (QZ-XM), Sanya (SY), and one outgroup individual from Thailand (TH). The phylogenetic tree based on SNPs from whole genome

resequencing shows that these populations have a clear genetic structure corresponds to their geographic distribution. However, population genetic analysis based on the MHC region does not distinctly separate the five subpopulations. Details are shown in Figure 6.

The principal component analysis (PCA) of the whole genome resequencing data in *S. chinensis* reveals obvious differentiation between the individual from Thai waters and the South China Sea populations. The South China Sea population can be roughly divided into four subgroups: LZB, SNB, QZ-XM, and SY (Figures 6C & E). Our results clearly show that though different *S. chinensis* populations have clear genetic structure at genomic level, the MHC genes seem not play a key role for their population genetic differentiation.

Population-level features of MHC environmental adaptation

The Pi results for the *S. chinensis* populations show that the QZ-XM population has higher values compared to the LZB and SNB populations. The polymorphism of MHC class IIb and some class I gene loci is higher than that of other regions (Figure 7 & Table S13).

Several gene loci, including *DQB*, *OOEP*, *DQA*, *DRB*, *BTN1A1* from class II, *BTN3A3*, *TRIM26*, *TRIM40* from class I, and *CFB*, *SFTA2* from class III, exhibit higher (top 5%) polymorphism compared to other MHC genes (Figure 7). Immune-related genes are typically crucial for individual survival.^{58–60} We further conducted functional enrichment analysis on these genes with high genetic polymorphism. We find that they are significantly enriched in the “positive regulation of immune response” pathway ($P < 0.05$). These gene regions may provide the population with a defensive advantage against pathogens through genetic diversity. We also find these genes enriched in “interferon gamma signaling” pathway, which plays a key role in antiviral responses and immune system regulation. High diversity in this pathway could be the result of the population adapting to various viruses or other pathogens (Figure S6).

Tajima's *D* analysis shows that most MHC genes have a low level of selection. However, some gene loci with Tajima's *D* obviously greater than 0 (top 1%) suggest that these genes might be under higher levels of balancing selection (Figure 7 & Table S11). Most of the selected genes are from LZB with a small proportion from QZ-XM and SNB, indicating that LZB may suffer population contraction now. The strong selected genes in LZB population are primarily enriched in the “positive regulation of immune response” pathway, suggesting that balancing selection has likely maintained the genetic diversity of pathogen resistance genes to cope with a diverse infection environment (Figure S7A). Certain gene loci with Tajima's *D* obviously less than 0 (least 1%) suggest these loci may have undergone a selective sweep, likely due to the rapid spread of advantageous mutations within the population. This is particularly evident for some loci in SNB. Functional enrichment analysis of these genes shows that they are mainly enriched in pathways related to positive regulation of immune response, and positive regulation of DNA-binding transcription factor (Figure S7B).

Based on the combined results of nucleotide diversity (Pi) and Tajima's *D* analyses, a total of nine genes, i.e., *TRIM26*, *TRIM40*, *BTN3A3* from class I, *OOEP*, *DQB*, *DQA*, *DRB*, *BTN1A1* from class II, *CFB* from class III exhibit both high genetic polymorphism and strong signatures of selection. These findings suggest that most of the genes with high levels of polymorphism (9/10) encountered strong natural selection.

By calculating the genetic differentiation index (*Fst*) between pairwise combinations of the *S. chinensis* populations (Table S14 & Figure S8), we observe varying degrees of MHC differentiation between different populations and regions. As shown in Figure S8, there is substantial differentiation mainly occurs in the MHC class I and II regions. Specifically, the differences between the LZB and SNB populations mainly occur in the I-extended region, ($Fst > 0.25$, Figure S8A); while LZB and QZ-XM in the front portions of the IIa, IIb, and I regions (Figure S8B). The specific loci showing differentiation are provided in Table S12.

Based on the top 5% of values from *Fst* (threshold $Fst > 0.30$) and XP-CLR (threshold XP-CLR > 1.74) analyses, four genes are identified as candidates under strong natural selection in different populations of the *S. chinensis*. They are *DOB*, *DMA*, *BTN1A1* from class II, and *NOTCH4* from class III.

Integrating population genetic analyses (Pi, Tajima's *D*, *Fst* and XP-CLR), we find that *BTN1A1* may play critical roles in the adaptation of *S. chinensis*

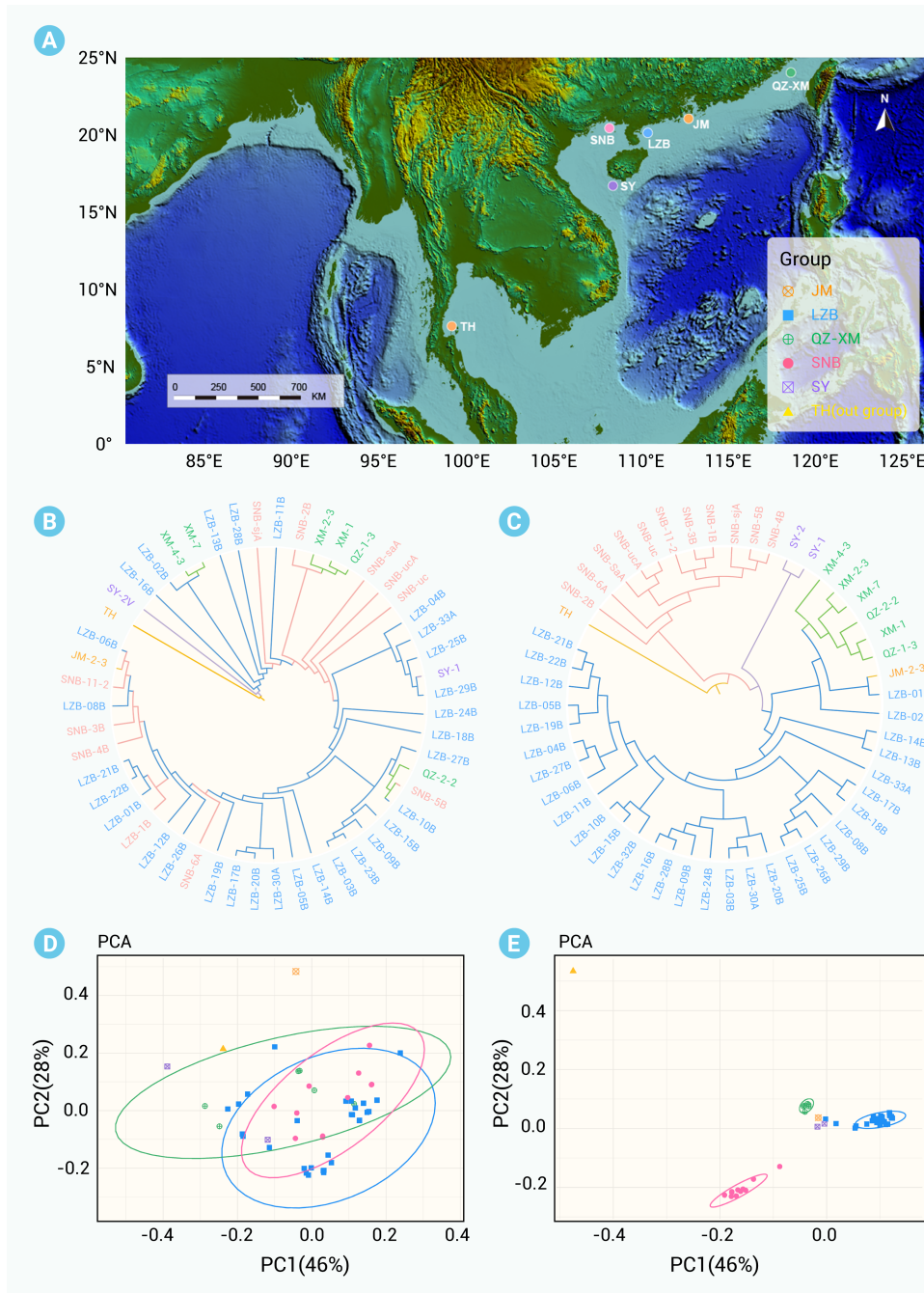


Figure 6. Geographic distribution and phylogenetic relationships of six populations of the *S. chinensis* (A) depicts the geographic distribution of the *S. chinensis* samples collected in this study. (B) and (C) show the phylogenetic trees based on the MHC region and whole-genome resequencing data, respectively. (D) and (E) display the principal component analysis (PCA) of the MHC region and whole-genome resequencing data, respectively.

CONCLUSIONS

In this study, we identified and isolated the complete MHC regions of four representative cetacean species and compared them with their terrestrial relatives to reveal adaptation strategies for water environment immunity throughout cetacean evolutionary history.

In cetaceans, the proportion of repetitive sequences in the MHC region slightly decreases, indicating a streamlined or specialized MHC function compared to terrestrial relatives. RTE-BovB elements are nearly absent, and LTR elements are significantly reduced in both MHC and the whole genome ($P < 0.05$), likely as an adaptation to aquatic environments. L1/CIN4 elements are expanded, with higher proportions in cetacean genomes than in MHC genes, suggesting positive adaptation to aquatic conditions but limited relevance to immune adaptation. Cetaceans exhibit significantly higher proportions of SINE elements, more concentrated in MHC genes than in genomes, indicating their importance for immunity and aquatic adaptation.

MHC regions in cetaceans appear more compact, with reduced gene content, particularly in extended class I and class II regions, reflecting a potentially streamlined immune system adapted to aquatic environments. Despite gene loss or pseudogenization, cetaceans maintain basic synteny and framework regions. class III genes show the highest conservation (43.3% of conserved orthologous genes), while class II genes exhibit the lowest, indicating greater variability in class I and II genes. MHC gene density in cetaceans is significantly higher (50–60 genes/Mb), primarily due to elevated density in the class II region (30–42

populations to their specialized habitats. As a member of the butyrophilin (*BTN*) gene family, *BTN* is genetically associated with inflammatory diseases and are predominately located within the MHC, *BTN* proteins function in T cell inhibition and epithelial cell–T cell interactions, suggesting their involvement in fine-tuning immune responses to local environmental pressures.⁶¹

Our analyses also reveal that the *TRIM* gene family, particularly *TRIM26*, and *TRIM40*, experienced positive selection during the transition from terrestrial to aquatic environments, followed by further diversification across ecological niches. High interpopulation diversity (P_i , F_{st}) and strong selection signals (Tajima's D , XP-CLR) indicate that these genes were also very important for *S. chinensis* populations get accustomed to their special habitats. *TRIM* proteins are known regulators of autophagy, immunity, and carcinogenesis.⁶² Previous studies have shown that *TRIM26* has been shown to participate in various physiological and pathological processes, including involvement in numerous malignancies and psychiatric disorders.⁶³ Given their roles as described above, future studies combining environmental genomics and functional assays could elucidate the precise mechanisms by which these genes facilitate habitat-specific adaptation in cetaceans.

genes/Mb) compared to terrestrial mammals (20–25 genes/Mb); other regions have comparable densities.

The expanded gene set in cetaceans enriches terms including negative regulation of nitrogen compound metabolism (GO:0051172), macromolecule biosynthetic processes (GO:0010558), and cellular metabolic processes (GO:0031323), potentially linked to low-oxygen adaptation. In *S. chinensis*, expanded genes enrich pathways for necroptosis (ko04217) and neutrophil extracellular trap formation (ko04613), possibly adapting to coastal pathogen pressures. Contracted genes in cetaceans include olfactory transduction (ko04740), associated with aquatic adaptations, while in *S. chinensis*, they relate to reproduction, suggesting lower immune pressure in coastal breeding.

Phylogenetic analyses indicate that MHC class I and II genes are vital for cetacean water environment adaptation, with class II genes critical for ancestral rapid adaptation and class I genes for subsequent habitat diversification. Highly conserved class III genes show little correlation with aquatic adaptation. Species-level selection analyses reveal strong positive selection for *DRA* (class II), *LTB* (class III), and *ZBED9*, *C6orf15*, *MRPS18B*, *MOG*, *TRIM15*, *PPP1R11* (class I) throughout cetacean evolution, crucial for aquatic adapta-

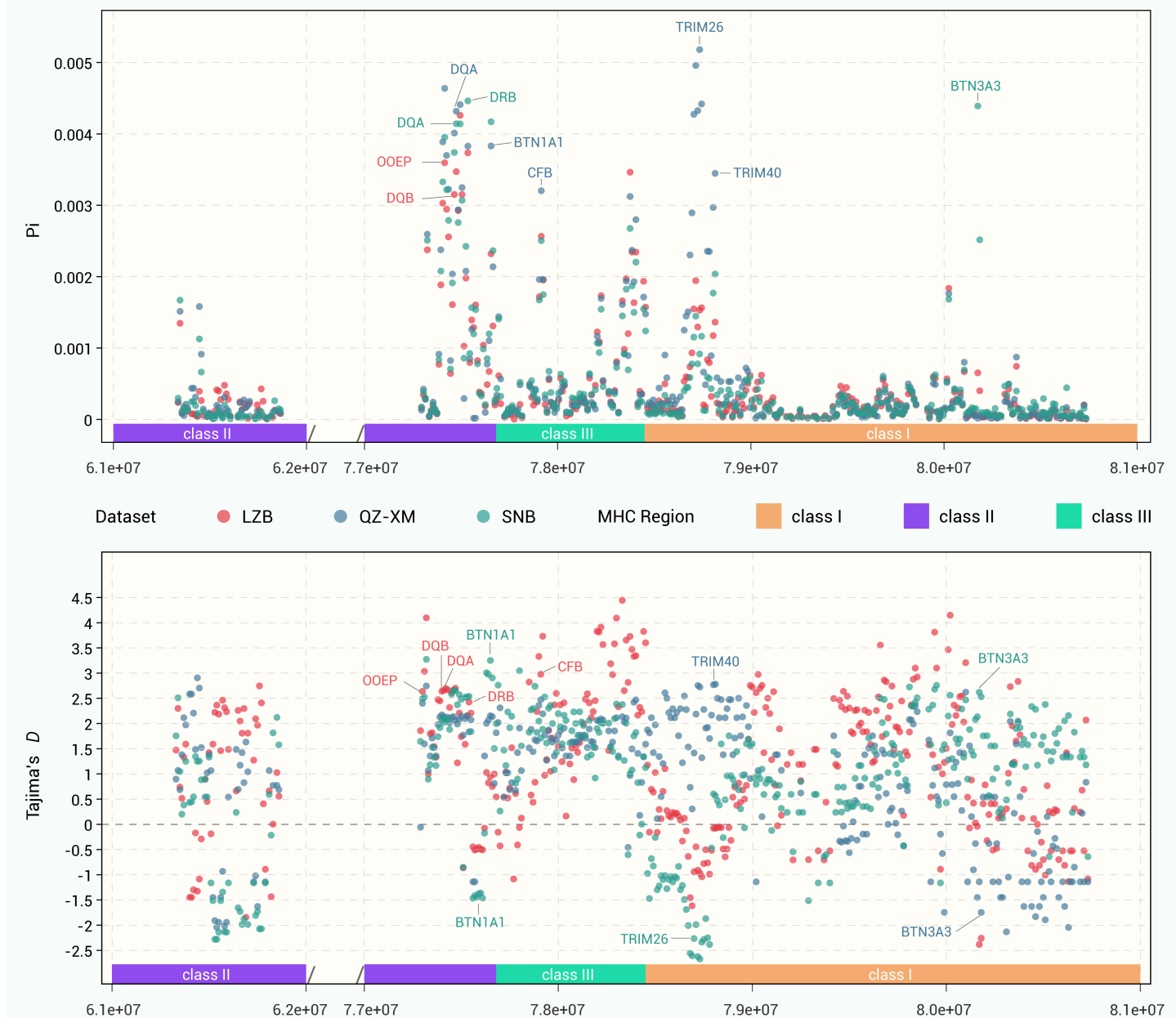


Figure 7. Polymorphism analysis of Pi and Tajima's D in *S. chinensis* populations.

tion and *S. chinensis*'s nearshore success. Population-level analyses identify natural selection for *DOB*, *DMA*, *BTN1A1* (class II), and *NOTCH4* (class III) across *S. chinensis* populations, with 9 genes, i.e., *TRIM26*, *TRIM40*, *BTN3A3* (class I), *OOEP*, *DQB*, *DQA*, *DRB*, *BTN1A1* (class II), and *CFB* (class III), showing high polymorphism and strong selection signatures within different *S. chinensis* populations. Besides, the evolutionary patterns of MHC genes differ significantly from those observed at the genome-wide level between different *S. chinensis* populations, suggesting that previous studies utilizing MHC genes as genetic markers to infer population genetic structure should be interpreted with caution.

We would also like to emphasize that limited sample sizes may result in highly skewed and statistically unreliable signals in population genetics analyses. To ensure the robustness of our conclusions, populations with insufficient sample sizes (i.e., SY, JM, and TH) have been excluded from inter-population statistical comparisons. Meanwhile, the QZ-XM population also exhibits a relatively small sample size compared to the LZB and SNB populations. Therefore, the primary findings on MHC gene selection at the population level may need to be validated in larger and more diverse populations. In

addition, while we employed GO and KEGG enrichment analyses to infer potential functional adaptations of strongly selected genes in this study, it is important to note that these bioinformatic predictions, based solely on genomic sequence data or gene set enrichment, cannot directly confirm physiological adaptations. The identified functional associations should be regarded as putative hypotheses. Thus, further experimental validation, such as transcriptomic profiling or functional assays, will be necessary to confirm the specific molecular mechanisms underlying these adaptations.

In summary, cetaceans exhibit streamlined MHC regions with reduced repetitive elements and gene content, yet maintain conserved structural frameworks. Key adaptations include loss of RTE-BovB elements, contracted LTR elements, expanded SINE elements and specific gene enrichments for low-oxygen tolerance, while class I and II genes drive evolutionary responses to aquatic habitats, supported by strong selection pressures across species and populations. Our study also showcases that marine ecosystems may be more stable and hospitable than terrestrial ones, thereby maintaining reduced selective pressures—this is why cetaceans appear to have streamlined immune gene repertoires.

REFERENCES

- Liu J., Lu Z., Wang X., et al. (2024). From land to sea: Unique evolution in reproductive strategies of marine mammals. *Innov. Life* **2**:100090. DOI:10.59717/j.xinn-life.2024.100090
- Zhang P., Goodman S. J., O'Connell M. J., et al. (2023). Marine mammal genomes: Important resources for unravelling adaptation and evolution in the marine environment. *Innov. Geosci.* **1**:100022. DOI:10.59717/j.xinn-geo.2023.100022
- Mössner S. and Ballschmiter K. (1997). Marine mammals as global pollution indicators for organochlorines. *Chemosphere* **34**:1285–1296. DOI:10.1016/s0045-6535(97)00426-8
- Lefebvre K. A., Charapata P., Stimmelmayr R., et al. (2025). Bowhead whale faeces link increasing algal toxins in the Arctic to ocean warming. *Nature* **644**:693–698. DOI:10.1038/s41586-025-09230-5
- Vermeij G. J. (2017). How the land became the locus of major evolutionary innovations. *Curr. Biol.* **27**:3178–3182. e3171. DOI: 10.1016/j.cub.2017.08.076
- Huelsmann M., Hecker N., Springer M. S., et al. (2019). Genes lost during the transition from land to water in cetaceans highlight genomic changes associated with aquatic adaptations. *Sci. Adv.* **5**:eaaw6671. DOI:10.1126/sciadv.aaw6671
- Marcovitz A., Turakhia Y., Chen H. I., et al. (2019). A functional enrichment test for molecular convergent evolution finds a clear protein-coding signal in echolocating bats and whales. *Proc. Natl. Acad. Sci. USA* **116**:21094–21103. DOI:10.1073/pnas.1818532116
- Yuan Y., Zhang Y., Zhang P., et al. (2021). Comparative genomics provides insights into the aquatic adaptations of mammals. *Proc. Natl. Acad. Sci. USA* **118**:e2106080118. DOI:10.1073/pnas.2106080118
- Xu S., Shan L., Tian R., et al. (2025). Multi-level genomic convergence of secondary aquatic adaptation in marine mammals. *The Innovation* **6**:100789. DOI:10.1016/j.xinn.2025.100798
- Shen T., Xu S., Wang X., et al. (2012). Adaptive evolution and functional constraint at TLR4 during the secondary aquatic adaptation and diversification of cetaceans. *BMC Evol. Biol.* **12**:39. DOI:10.1186/1471-2148-12-39
- Kelley J., Walter L. and Trowsdale J. (2005). Comparative genomics of major histocompatibility complexes. *Immunogenetics* **56**:683–695. DOI:10.1007/s00251-004-0717-7
- Benacerraf B. (1981). Role of MHC gene products in immune regulation. *Science* **212**:1229–1238. DOI:10.1126/science.6165083
- Radosavljevic M., Cuillerier B., Wilson M. J., et al. (2002). A cluster of ten novel MHC class I related genes on human chromosome 6q24.2-q25.3. *Genomics* **79**:114–123. DOI:10.1006/geno.2001.6673
- Trowsdale J. (2011). The MHC, disease and selection. *Immunol. Lett.* **137**:1–8. DOI:10.1016/j.imlet.2011.01.002
- Piertney S. B. and Oliver M. K. (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity* **96**:7–21. DOI:10.1038/sj.hdy.6800724
- Goldberg A. C. and Rizzo L. V. (2015). MHC structure and function – antigen presentation. Part 1. *Einstein-Sao Paulo* **13**:153–156. DOI:10.1590/s1679-45082015rb3122
- Bartl S., Baish M. A., Flajnik M. F., et al. (1997). Identification of class I genes in cartilaginous fish, the most ancient group of vertebrates displaying an adaptive immune response. *J. Immunol.* **159**:6097–6104. DOI:10.4049/jimmunol.159.12.6097
- Bartl S. and Weissman I. L. (1994). Isolation and characterization of major histocompatibility complex class II B genes from the nurse shark. *Proc. Natl. Acad. Sci. USA* **91**:262–266. DOI:10.1073/pnas.91.1.262
- Kasahara M., Vazquez M., Sato K., et al. (1992). Evolution of the major histocompatibility complex: Isolation of class II A cDNA clones from the cartilaginous fish. *Proc. Natl. Acad. Sci. USA* **89**:6688–6692. DOI:10.1073/pnas.89.15.6688
- Ohta Y., McKinney E. C., Criscitiello M. F., et al. (2002). Proteasome, transporter associated with antigen processing, and class I genes in the nurse shark *Ginglymostoma cirratum*: Evidence for a stable class I region and MHC haplotype lineages. *J. Immunol.* **168**:771–781. DOI:10.4049/jimmunol.168.2.771
- Okamura K., Ototake M., Nakanishi T., et al. (1997). The most primitive vertebrates with jaws possess highly polymorphic MHC class I genes comparable to those of humans. *Immunity* **7**:777–790. DOI:10.1016/s1074-7613(00)80396-9
- Belov K., Deakin J. E., Papenfuss A. T., et al. (2006). Reconstructing an ancestral mammalian immune supercomplex from a marsupial major histocompatibility complex. *PLoS Biol.* **4**:e46. DOI:10.1371/journal.pbio.0040046
- de Sá A. L. A., Breaux B., Burlamaqui T. C. T., et al. (2019). The marine mammal class II major histocompatibility complex organization. *Front. Immunol.* **10**:696. DOI:10.3389/fimmu.2019.00696
- Ruan R., Ruan J., Wan X. L., et al. (2016). Organization and characteristics of the major histocompatibility complex class II region in the Yangtze finless porpoise (*Neophocaena asiaeorientalis asiaeorientalis*). *Sci. Rep.* **6**:22471. DOI:10.1038/srep22471
- Reche P. A. and Reinherz E. L. (2003). Sequence variability analysis of human class I and class II MHC molecules: Functional and structural correlates of amino acid polymorphisms. *J. Mol. Biol.* **331**:623–641. DOI:10.1016/s0022-2836(03)00750-2
- Spurgin L. G. and Richardson D. S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc. Biol. Sci.* **277**:979–988. DOI:10.1098/rspb.2009.2084
- Nei M., Gu X. and Sitnikova T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. USA* **94**:7799–7806. DOI:10.1073/pnas.94.15.7799
- Jefferson T. A., Webber M. A. and Pitman R. L. (2015). Marine mammals of the world (Second Edition) (Academic Press). DOI: 10.1016/B978-0-12-409542-7.50004-4
- Xu S., Zhang P., Ij S., et al. (2009). A preliminary analysis of genetic variation at three MHC loci of the Indo-Pacific humpback dolphin (*Sousa chinensis*). *Acta Theriologica Sinica* **29**:372–381. <https://www.mammal.cn/EN/Y2009/V29/I4/372>
- Zhang P., Zhao Y., Li C., et al. (2020). An Indo-Pacific humpback dolphin genome reveals insights into chromosome evolution and the demography of a vulnerable species. *iScience* **23**:101640. DOI:10.1016/j.isci.2020.101640
- Flynn J. M., Hubley R., Goubert C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA* **117**:9451–9457. DOI:10.1073/pnas.1921046117
- Ou S. and Jiang N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**:1410–1422. DOI:10.1104/pp.17.01310
- Chen N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **5**:4–10. DOI:10.1002/0471250953.bi0410s05
- Stanke M., Keller O., Gunduz I., et al. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**:W435–439. DOI:10.1093/nar/gkl200
- Binney E., Clamp M. and Durbin R. (2004). GeneWise and Genomewise. *Genome Res.* **14**:988–995. DOI:10.1101/gr.1865504
- Haas B. J., Salzberg S. L., Zhu W., et al. (2008). Automated eukaryotic gene structure annotation using EVIDENCEModeler and the program to assemble spliced alignments. *Genome Biol.* **9**:R7. DOI:10.1186/gb-2008-9-1-r7
- Buchfink B., Xie C. and Huson D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**:59–60. DOI:10.1038/nmeth.3176
- Wang Y., Tang H., Debarry J. D., et al. (2012). MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**:e49. DOI:10.1093/nar/gkr1293
- Emms D. M. and Kelly S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**:238. DOI:10.1186/s13059-019-1832-y
- Altschul S. F., Gish W., Miller W., et al. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410. DOI:10.1016/s0022-2836(05)80360-2
- Rokas A. (2011). Phylogenetic analysis of protein sequence data using the Randomized Accelerated Maximum Likelihood (RAXML) Program. *Curr. Protoc. Mol. Biol.* **96**:19.11.1–19.11.14. DOI:10.1002/0471142727.mb1911s96
- Yang Z. and Rannala B. (2006). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**:212–226. DOI:10.1093/molbev/msj024
- Yang Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**:1586–1591. DOI:10.1093/molbev/msm088
- Horton R., Wilming L., Rand V., et al. (2004). Gene map of the extended human MHC. *Nat. Rev. Genet.* **5**:889–899. DOI:10.1038/nrg1489
- Fukami-Kobayashi K., Shiina T., Anzai T., et al. (2005). Genomic evolution of MHC class I region in primates. *Proc. Natl. Acad. Sci. USA* **102**:9230–9234. DOI:10.1073/pnas.0500770102
- Martelossi J., Nicolini F., Subacchi S., et al. (2023). Multiple and diversified transposon lineages contribute to early and recent bivalve genome evolution. *BMC Biol.* **21**:145. DOI:10.1186/s12915-023-01632-z
- Wong L. H. and Tremethick D. J. (2024). Multifunctional histone variants in genome function. *Nat. Rev. Genet.* **26**:82–104. DOI:10.1038/s41576-024-00759-1
- Cassandri M., Smirnov A., Novelli F., et al. (2017). Zinc-finger proteins in health and disease. *Cell Death Discov.* **3**:17071. DOI:10.1038/cddiscovery.2017.71
- Rogers L. W., Koehler A. M., Crouch J. A., et al. (2022). Comparative genomic analysis reveals contraction of gene families with putative roles in pathogenesis in the fungal boxwood pathogens *Calonectria henricotiae* and *C. pseudonaviculata*. *BMC Ecol. Evol.* **22**:79. DOI:10.1186/s12862-022-02035-4
- Demuth J. P. and Hahn M. W. (2009). The life and death of gene families. *Bioessays* **31**:29–39. DOI:10.1002/bies.080085
- Freitas L. and Nery M. F. (2020). Expansions and contractions in gene families of independently-evolved blood-feeding insects. *BMC Evol. Biol.* **20**:87. DOI:10.1186/s12862-020-01650-3
- Fabrizius A., Hoff M. L., Engler G., et al. (2016). When the brain goes diving: Transcriptome analysis reveals a reduced aerobic energy metabolism and increased stress proteins in the seal brain. *BMC Genomics* **17**:583. DOI:10.1186/s12864-016-2892-y
- Mazzatenta A., Bondi D., Di Giulio C., et al. (2022). Olfactory response to altitude hypoxia: A pilot study during a Himalayan trek. *Adv. Exp. Med. Biol.* **1375**:55–61. DOI:10.1007/5584_2021_679
- Lu T. (2023). The High-altitude adaptation mechanism of Vulpes in the Qinghai-Tibet Plateau. (Northeast Forestry University). DOI: 10.27009/d.cnki.gdblu.2023.000021
- Kishida T., Thewissen J., Hayakawa T., et al. (2015). Aquatic adaptation and the evolution of smell and taste in whales. *Zoological Lett.* **1**:9. DOI:10.1186/s40851-014-0002-z
- Browning J. L., Ngam-ek A., Lawton P., et al. (1993). Lymphotoxin beta, a novel member of the TNF family that forms a heteromeric complex with lymphotoxin on

- the cell surface. *Cell* **72**:847–856. DOI:10.1016/0092-8674(93)90574-a
57. Browning J. L., Androlewicz M. J. and Ware C. F. (1991). Lymphotoxin and an associated 33-kDa glycoprotein are expressed on the surface of an activated human T cell hybridoma. *J. Immunol.* **147**:1230–1237. DOI:10.4049/jimmunol.147.4.1230
 58. Dahmer M. K., Cornell T. and Quasney M. W. (2016). Genetic and epigenetic factors in the regulation of the immune response. *Curr. Opin. Pediatr.* **28**:281–286. DOI:10.1097/mop.0000000000000356
 59. Nguyen K. H., Le N. V., Nguyen P. H., et al. (2025). Human immune system: Exploring diversity across individuals and populations. *Heliyon* **11**:e41836. DOI:10.1016/j.heliyon.2025.e41836
 60. Mehta S. and Jeffrey K. L. (2015). Beyond receptors and signaling: Epigenetic factors in the regulation of innate immunity. *Immunol. Cell Biol.* **93**:233–244. DOI:10.1038/icb.2014.101
 61. Abeler-Dörner L., Swamy M., Williams G., et al. (2012). Butyrophilins: An emerging family of immune regulators. *Trends Immunol.* **33**:34–41. DOI:10.1016/j.it.2011.09.007
 62. Hatakeyama S. (2017). TRIM family proteins: Roles in autophagy, immunity, and carcinogenesis. *Trends Biochem. Sci.* **42**:297–311. DOI:https://doi.org/10.1016/j.tibs.2017.01.002
 63. Zou J., Niu K., Lu T., et al. (2024). The multifunction of TRIM26: From immune regulation to oncology. *Protein Pept. Lett.* **31**:424–436. DOI:10.2174/0109298665311516240621114519

FUNDING AND ACKNOWLEDGMENTS

This work is supported by Innovation Project for Science and Technology Talents from Hainan Provincial Natural Science Foundation of China (KJRC2023C36), and the Fujian

Provincial Natural Science Foundation (2023J01210). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization: Peijun Zhang; Methodology: Dongjiao Liu, Jun Liu, Peiyu Xie, Kui Kang, Peijun Zhang; Visualization: Dongjiao Liu, Jun Liu; Supervision: Peijun Zhang, Xianyan Wang; Writing-original draft: Dongjiao Liu; Writing-review & editing: Peijun Zhang, Songhai Li, Xianyan Wang, Liyuan Zhao, Jun Liu, Qianhui Zeng. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

ETHICAL STATEMENT AND PATIENT CONSENT

Not applicable.

DATA AND CODE AVAILABILITY

Data are available from the corresponding author upon reasonable request.

SUPPLEMENTAL INFORMATION

It can be found online at <https://doi.org/10.59717/j.xinn-life.2026.100245>